

Report on Validity Issues in *[redacted]*

Glenn Beamer, Rutgers University
July 7, 2005

Report on Validity and Reliability Issues in [redacted]

Glenn Beamer, Rutgers University
July 7, 2005

Executive Summary

1. Survey and reporting forms should provide valid and reliable information that reflects past events and respondent characteristics. In this case validity has two components – content validity and measurement validity. Content validity is the extent to which individual queries measure characteristics and events such that they can be readily related to a relevant dimension of an employment history. Measurement validity is the degree to which individual queries accurately, systematically, and dependably report the extent of characteristics and frequency of events. Reliability is the extent to which individual queries are likely to be reported accurately by numerous recorders taking into account human error and perceptual differences that can contribute to varied recordations. By maximizing the extent to which any two recorders will accurately report the same events and characteristics, regardless of other factors external to the survey, researchers minimize the unreliability of survey data.
2. Content validity requires that individual queries accurately measure individual events and characteristics related to a specific dimension of an employee's character or work history. With respect to employment histories, for a survey to have construct validity questions about individual events and characteristics should be connected to underlying dimensions of employment history such as safety, customer relations, trustworthiness, timeliness, and policy compliance. The flawed design of the [redacted] Termination Record form, its lack of topical organizations, and the lack of connections among topics and individual measures eliminate clarity from the termination record form work record section and render it invalid with respect to its contents.
3. Measurement validity is the extent to which individual queries accurately portray the frequency of an event or the extent of a characteristic. In the [redacted] Termination Record form work record section the biased and inadequate scale used for 17 out of 20 individual measures as well as ambiguous instructions and definitions render the measurements invalid.
4. Reliability is the extent to which recorders succeed or fail at accurately representing, literally re-presenting, an employee's character and work events. Whereas measurement validity pertains to how well or poorly individual queries reflect event frequency or the extent of character traits, reliability results from respondents consistently and accurately portraying events and characteristics. Unreliable data can result from human error, from recorders having inconsistent understandings of how and when to record events and characteristics (i.e. ambiguous or uninformative survey directions), and from recorders inconsistently using different categorical responses to reflect the same event or characteristic (i.e. inconsistent application of the survey directions). The analysis of the data from a sample of 5000 [redacted] Termination Record forms and from work record code user profiles of 3603 carriers reveals that driver characteristics and employment events are inconsistently and inaccurately reported and that [redacted]

Termination Record forms fail to provide reliable representations of drivers' characteristics and work histories.

5. The definitions and directions provided by [redacted] for recorders using the [redacted] Termination Record form provide inadequate guidance and contribute to the unreliability of the data produced by carriers.
6. Data generated by [redacted] Termination Record forms suffers from the lack of content validity, invalid measurements, and unreliable reporting. Consequently:
 - a. [redacted] and carriers using the [redacted] Termination Record form have no means by which to organize, aggregate, or index data such that they can make meaningful inferences about an individual driver's professional characteristics or event history (i.e. the [redacted] Termination Record form lacks content validity).
 - b. Carriers using [redacted] Termination Record forms can have no confidence that they are assessing an employee's work history as it pertains to the frequency of events or the extent of characteristics (i.e. the [redacted] Termination Record form lacks measurement validity).
 - c. Carriers using the [redacted] Termination Record form can have no confidence that recorders have consistently and accurately reported drivers' characteristics and employment histories (i.e. the data produced by [redacted] Termination Record forms are unreliable).
7. Statistical analyses of [redacted] Termination Record forms reveal that hundreds of carriers use individual codes in ways that are statistically deviant from ranges expected in varied samples and that indicate systemic distortions. Carriers used codes at rates 4 to 33 times greater than would have been expected in the absence of systemic distortions.
8. A reliability analysis of [redacted] Termination Record form work record codes produces a reliability coefficient of only 0.157 on a 0 to 1.0 scale. Reliable data typically have corresponding coefficients in excess of 0.7 if not 0.8. The estimated error rate for [redacted] Termination Record forms is 60%.

Introduction: Content Validity, Measurement Validity, and Reliability

Content validity, measurement validity, and reliability are three qualities of survey research relevant to the case of *[redacted]*. The survey is the *[redacted]* Termination Record form and the measures are the individual queries designed to record events and their frequency and driver characteristics and their extent.

This report begins with a section in which I define content validity and measurement validity. I explain content validity, measurement validity, and their pertinence to the *[redacted]* Termination Record form. Section 2 provides an analysis of the content and measurement validity of the *[redacted]* Termination Record form's work record queries. I rely upon the termination record form and its instructions to carriers regarding how to record events and driver characteristics. In section 3 I evaluate the termination record form based upon the data produced by carriers who have used the *[redacted]* Termination Record form. In other words section 2 presents an *ex ante* evaluation of the form as it is presented to carriers and section 3 presents an *ex post* evaluation in which I assess employment histories produced by carriers using the *[redacted]* Termination Record form.

The evaluation in section 2 establishes that *[redacted]* fails to have content validity in the work record section of the termination record form and that the individual work record measures are inadequate and biased. Because the measures' scales are inadequate and biased the measures themselves are invalid. The analysis in section 3 reveals that carriers behave aberrantly when using work record codes and that carriers do not share a common understanding of how and when to use different codes.

Section 4 presents a basic analysis of the termination record form work record section and a statistical analysis focusing on the reliability of the employment histories produced by the work record section. The basic analysis reveals that at least ten percent of termination record forms are incomplete and cannot be relied upon as meaningful work histories. The statistical analysis reveals that carriers use the work record section inconsistently and the data produced to represent employment events and driver characteristics are unreliable. I estimate an error rate for the work record section of 60 percent.

Section 1: Content Validity and Measurement Validity

Content validity and measurement validity are qualities of empirical reports such as employment histories produced from survey forms such as employee evaluations. Content validity refers to how clearly individual survey measures relate to topics relevant to the research subject. Measurement validity regards the extent to which survey queries accurately gauge and denote individual items relevant to various topics. Content validity and measurement validity are not absolutes (Carmines and Zeller, 1978; Litwin, 1995). Research instruments are evaluated as having more or less content validity and more or less measurement validity with respect to the individual measures, topical categories, and overall subject (Babbie, 2004). In [redacted], the research instrument is the [redacted] Termination Record form and the data it produces are drivers' employment histories. The [redacted] Termination Record form's subject is the employment history of an individual driver. Categorical topics within an employment history could include safety, professional standards and conduct, and customer relations (U.S. Department of Labor, 2005). Individual queries are the measures that comprise a driver's profile for each topic and these topics then comprise an employment history.

Evaluating content validity is a qualitative exercise that is nevertheless rigorous and thorough (Babbie, 1994). By examining the organization and evaluating the clarity of a survey *Termination Record form* such as the termination record form, one can assess its content validity. Both content and measurement validity can be evaluated by comparison with similar surveys (Litwin, 1995; Fowler, 1995)). Attached as Appendix A by way of illustration are the termination record form, two driver evaluation forms used from public school districts. Attached as Appendix B by way of illustration is an alternative work record section of the termination record form that I composed. I compare the termination record form to these employment history reports. These comparisons lead to specific changes that if made would increase substantially the [redacted] Termination Record form's validity.

Evaluating Content Validity and Measurement Validity. There are three criteria that we can use to assess the content and measurement validity – clarity, scale, and bias (Babbie, 2004; Judd, Kidder, and Smith, 1991; Litwin, 1995; Carmines and Zeller, 1978). Clarity relates directly to content validity (Babbie, 2004). Scale and bias relate to measurement validity (Litwin, 2004; Fowler, 1995). I discuss each of these in turn.

Content Validity & Clarity. In order for surveys to be meaningful, the individual measures and topics researched should be clearly identified for both the research producer and

consumer (Babbie, 2004). In [redacted], the research producers are previous employers who submit [redacted] Termination Record forms to [redacted]. [redacted] in turn uses submitted termination record forms to produce employment histories. The research consumers are prospective employers who evaluate the employment histories. Clear definitions and categorical organization increase the content validity of surveys by enhancing their straightforward production and interpretation (Babbie, 2004).

Content validity involves identifying and using appropriate measures as components of topics that are relevant to the research subject (Babbie, 2004; Litwin, 1995). The [redacted] Termination Record form is an instrument that carriers use and which [redacted] collects to produce employment histories. As a research instrument the [redacted] Termination Record form should include measures of dimensions of truck drivers' employment histories. Thus the subject of the research being conducted by a prospective employer is the driver's employment history. Three topics considered relevant to that research subject, the employment history, would likely include driver safety practices, driver professional conduct, and driver customer relations. Individual measures of driver safety could include measures of events such as accidents, measures of compliance with safety regulations posted by the company and law enforcement, and measures of third party observations such as merit citations or records of complaints from drivers or coworkers. Measures of professional conduct could include attendance and timeliness, schedule maintenance, relations with coworkers, relations with supervisors, and communication records.

The first step toward establishing clarity is to identify the relevant topics and the measures related to those topics for producers and consumers. Typically this step involves organizing measures in groups that correspond with a topic. In situations in which a measure could apply to more than one topic, clear organization can direct information producers and consumers as to which topic or topics the measure is intended to relate (Litwin, 1995).

A second step in establishing content validity is to organize a survey by categorical topics (Fowler, 1995). For employment histories topical organization should enhance both content and measurement validity by focusing recorders on an employee's experiences and characteristics for each area of investigation.

To understand content validity an example is helpful. Suppose we have a classroom with twenty children and we are interested in their size. Size is a broad term – you might think of weight and I might think of height. Both height and weight are valid components of child's

size. Just as safety and professional conduct are valid topics for an employment history, size and weight are valid topics if our research subject is "childrens' size." If a third person came to our classroom with an eye chart and stated that she was going to test children's eyesight, we would understand that eyesight or visual impairment is not a topic related to size. Eyesight is a dimension of the subject "sensory ability," along with feeling, hearing, and taste. Height and weight are clearly dimensions of size and can be assessed as having content validity with respect to the subject size. Eyesight is an invalid dimension of size but a valid dimension of "senses."

Content validity also refers to the extent to which measures comprehensively cover a dimension of the research subject (Babbie, 2004, pp. 143-146). Just as height is only one dimension of size, an employment history that focused only on schedule maintenance would be lacking. In order to have an employment history with content validity other dimensions such as safety and customer relations should be included.

Measurement Validity, Scale, and Bias. Measurement validity refers to the extent to which an instrument or query accurately and fairly represents the frequency of an event or the extent of a characteristic (Fowler, 1995; Judd, Kidder, and Smith, 1991). Valid measures reflect the range of possible outcomes for queries relevant to a topic (Babbie, 2004; Carmines and Zeller, 1978). Valid measures should provide sufficient information such that research consumers, in [redacted] potential employers, understand that the information accurately reflects the extent of characteristics, such as children's heights, and the frequency of events.

Standard scales enhance measurement validity and help ensure that recorders understand and can use a range of available responses that correspond to the range of events and characteristics (Fowler, 1995). Survey measure scales should have two characteristics – they should be adequate and they should be consistent (Fowler, 1995). Adequate scales are appropriate for measuring the size, extent, and frequency of events and characteristics. This means that children's weight measurements would be made in pounds and ounces and not tons. Freight may be appropriately measured in tons and not ounces or milligrams.

When categorical responses, such as not satisfactory, satisfactory, and excellent are used in measurements, recorders should be given directions about the thresholds for using the different ordinal categories (Fowler, 1995; Litwin, 1995). Thresholds and guidelines for recording categorical responses enhance the consistency with which researchers use measurement scales. For example a researcher interested in employees' safety habits might

direct recorders to identify unsafe employees as those having "two or more accidents resulting in physical injury to themselves or others within the last year." This direction has three important components: 1) an empirical threshold of two or more accidents, 2) consequences of the accident which are physical injuries to the employee or others, and 3) a time frame that reflect the period for evaluation – the last year (Fowler, 1995).

In addition to providing responses for the range of outcomes adequate scales may reflect the context in which a given measure is taken. If we are measuring the frequency of school absences, we may want some indication of the number of days missed due to excused illnesses compared to unexplained absence. If we are measuring corporate performance, we may want to measure profits from operations and profits from asset sales. Conflating these two profit measures might provide a valid measure of "total profitability," but it would be incomplete and potentially misleading if we were looking to invest in the company.

In assessing adequacy we should ask "Are the measures adequate for the task at hand?" If we are designing a playhouse for a classroom full of children then knowing their heights and weights is adequate to build a structure in which the children won't get stuck, hit their heads, or fall through the platform. If we are sewing clothing for our classroom of children, then knowing their height and weight is probably not adequate for making each student a shirt. To sew shirts we would want to know the children's neck circumferences, their shoulder widths and their sleeve lengths. Adequate measures reflect the range of outcomes including the context in which events occurred or characteristics have been exhibited.

Unbiased. Good measures do not motivate recorders to classify subjects inappropriately. In addition to providing sufficient categories (adequacy) measures should not be subjective or should not focus on only negative characteristics and events or only on positive characteristics and events. Whereas comprehensive response categories and appropriate units of measurement contribute to adequacy, balanced identification and question wording create unbiased measures (Fowler, 1995).

Unbiased measures identify neither negative nor positive events and characteristics exclusively. Unbiased questions identify and record different types of events and characteristics and the frequency with which they occurred or the extent to which they were evident during employment histories. Unbiased questions should allow research consumers to discern, without great work, the extent of a characteristic relative to other survey respondents or the relative frequency of events. Biased questions often eliminate large subsets of respondents (in this

case drivers) from measurement because the questions do not apply or provide for inadequate responses that allow only negative inferences. Unbiased questions and responses provide the research consumer with information that can lead to either positive or negative inferences (Fowler, 1995).

To better understand measurement validity consider again a classroom of 20 children ages six to ten and for whom we want to record their heights. If a researcher came to the classroom with a bathroom scale, she could not record students' heights; she could record their weights. The bathroom scale is an invalid measurement instrument for recording height. Rulers, yardsticks, and a tape measure would all be valid instruments for measuring height. However, if a researcher brought a yardstick that was 40 inches long and claimed it measured three feet this would not be a valid research instrument. This 40-inch yardstick underreports children's heights by four inches per every three feet and its resulting measures are invalid. A yardstick is appropriate for measuring children's height and its scale, feet and inches, is both appropriate and adequate. But a 40-inch yardstick is a biased measurement instrument and thus yields biased and invalid measurements.

Section 2: Content Validity, Measurement Validity, & the [redacted] Form

For the analysis of content validity and measurement validity in the [redacted] Termination Record form I focus first on the form and its instructions. Neither the [redacted] Termination Record form nor its instructions establishes content validity because neither clearly identifies relevant categories of driver experiences and character.

The [redacted] Termination Record form's "work record" section has two categorical areas – work history and load abandonment. The first topic is denoted by work codes 901 through 938 and codes 957 through 999. The instructions include guidance for code 915, which no longer exists on the termination record form. Although [redacted] references "categories" for carriers to identify drivers as satisfactory, superior, or outstanding, [redacted] does not explicitly identify these categories on the instructions. For drivers to be recorded as satisfactory, superior, or outstanding [redacted] instructs employers that such drivers must meet or exceed individual company standards "in all categories." Presumably these are performance categories such as employee relations, customer relations, and safety but [redacted] makes no effort to identify the relevant categories nor does [redacted] specify which individual measures relate to these unidentified categories. Therefore there is no means by which to establish content validity for the Work Record section of the [redacted] Termination Record form when it is considered as a whole survey instrument. The failure to establish content validity rests squarely with [redacted] and [redacted].

For comparative evaluations I performed a "Google" search for "Driver Evaluation Forms." I printed forms from the University of Alabama Annual Employee Performance Evaluation, the Tacoma Public Schools Bus Driver Evaluation, the Rockingham County (Virginia) Public Schools Bus Driver Evaluation, and the State of Idaho (Bus) Driver Evaluation forms. All of these forms have identified work history categories each category has multiple measures. This search, form review, and printing required approximately forty minutes. The Tacoma and Rockingham County forms are reproduced in Appendix A.

The Tacoma and Rockingham driver evaluation forms have two features that provide greater content and measurement validity than the [redacted] Termination Record form work record section. Their content validity is established because the forms are organized into three or four work history categories such as professional conduct and interpersonal relations. The forms then have subordinate measures that clearly relate to these categories. The Tacoma

Public Schools forms organizes driver characteristics and employment events into 3 categories with 26 measures, the Rockingham County Schools form organizes driver characteristics and employment events into four categories with 30 measures. Taken together the topical profiles provide comprehensive reflections of bus driver characteristics and employment events.

The Tacoma and Rockingham County forms' organization provide clear direction for potential employers assessing drivers' work histories. In comparison the [redacted] Termination Record form conflates topical categories of employment history and compounds its disorganization by including identifiers that relate to the context of employment termination and requests for communications among carriers (codes 950 through 956 and code 944 respectively). Although [redacted] instructs previous employers to use all identifiers that apply, the vast majority of previous employers use only a single identifier. Table 1 compares the [redacted] Termination Record form with the Bus Driver Evaluation forms.

The Tacoma and Rockingham forms have standardized ordinal responses for each measurement. For each query recorders are asked to rate Rockingham drivers as 1) does not meet (the standard), 2) needs improvement, 3) satisfactory, 4) and outstanding. For each query on the Tacoma form evaluators are asked to rate drivers as 1) needs improvement, 2) meets expectations, 3) exceptional, 4) not observed. These comprehensive response scales permit evaluators to record an evaluation of every driver for every measure. In instances where the evaluator cannot reliably report an employment event or a driver characteristic the Tacoma form provides a response in which the evaluator can indicate his or her lack of observation. The Tacoma and Rockingham forms provide response scales that yield balanced and comparable measurements for all drivers. The [redacted] Termination Record forms provide neither evaluations of drivers for every measure nor balanced response codes that differentiate types of experiences.

**Table 2.1: Comparison of Work Record Forms
& Employment Histories**

Form	Identified Contents	Measures per Category	Format of Measures
Rockingham Public Schools Driver Evaluation Form	1) Organization & Planning 2) Interpersonal Relations & Communication 3) Professional Responsibilities/Qualities 4) Knowledge & performance of job responsibilities	5 4 7 14	Four category ordinal scales: 1) Outstanding, 2) Satisfactory, 3) Needs Improvement 4) Does Not Meet Drivers are evaluated with all 30 measures
Tacoma Public Schools Bus Driver Evaluation Form	1) Job Knowledge & Skills 2) Student Management Skills 3) Personal Qualities	12 4 10	Four category ordinal scale: 1) Exceptional 2) Meets Expectations 3) Needs Improvement 4) Not Observed Drivers are evaluated with all 26 measures.
<i>[redacted]</i> Form/Work Record Section 900	None identified but drivers must meet or exceed carrier expectations "in all categories" to be evaluated as satisfactory, superior, or outstanding.	20 work event and driver character measures plus 7 measures regarding location of termination.	Check if event occurred/characteristic applies. Carriers may circle as many as apply. Drivers are evaluated with zero to eighteen measures.

One potential objection to adopting a form similar to the Rockingham or Tacoma forms could be that these forms utilize an entire page and require employees' supervisors to record to thirty and twenty-six measures. Because of their organization, these forms require less consideration than the *[redacted]* Termination Record form. Supervisors using the school district forms understand that they record every measure for every driver. Carriers using the *[redacted]* Termination Record forms must decide which "identifiers" to use and which to

disregard. In cases where “identifiers” are not used, a potential employer cannot infer whether the identifier is unused because it does not apply or whether it is unused because the previous employer chose to ignore or skip the identifier. With the Rockingham and Tacoma forms unrecorded measures reflect incomplete employment histories. With the [redacted] Termination Record form unrecorded measures may or may not reflect incomplete employment histories. This point is relevant to the issue of data reliability as well as measurement validity. The superior measurement validity of the Rockingham and Tacoma forms enhances their reliability. The biased and inadequate scales in the termination record form likely compound their unreliability.

Fourteen of the twenty [redacted] work history measures ask previous employers to identify single, idiosyncratic events in a driver’s employment history.¹ These measures provide negative impressions about drivers who may have otherwise satisfactory, superior, or outstanding records. In contrast the Rockingham form uses a four category ordinal scale ranging from “Does Not Meet (Standards/Expectation)” to “Outstanding.” The Tacoma form uses a three category ordinal scale ranging from “Needs Improvement” to “Exceptional.” The Tacoma form also includes a category “Not observed.” This latter category enhances both measurement validity (events and characteristics that may not exist are not mistakenly represented) and the form’s reliability. The bus driver evaluations use qualifiers such as “As needed” to indicate whether drivers appropriately complied with school district policies. The [redacted] Termination Record form’s lack of such qualifiers deprives the measures of contextual representation and further undermines its measurement validity.

These fourteen measures deprive drivers of comprehensive employment histories. Identifying a driver as “polite and on time one day” would not provide a comprehensive employment history, and it is equally skewed and uninformative to use idiosyncratic negative events to reflect drivers’ experiences and characteristics. These measures are equivalent to using a yardstick to measure children’s heights but only measuring children who are less than three feet tall. Any children over three feet tall are left unrecorded and thus consumers of the children’s height research are left with an extremely biased representation of children’s heights.

¹ These fourteen identifiers are 913 – Cargo Loss, 915 – Falsified Employment Application, 917 – Equipment Loss, 924 – Late Pick Up/Delivery, 926 – Log Violation, 928 – No Show, 929 – Failed to Report Accident, 931 – Quit Under Dispatch 933 – Quit/Dismissed During Training and/or Orientation, 935 – Company Policy Violation, 957 – Unauthorized Equipment Use, 959 – Unauthorized Passenger, and 961 – Unauthorized Use of Company Funds.

Information consumers know only about relatively short children and have no means of knowing how *relatively* short they are because no data are recorded for taller children.

These same fourteen measures in the [redacted] Termination Record form inquire about only negative events such as "Late pick up," "unauthorized equipment use," or "cargo loss." These measures are biased. Unbiased measures would inquire about drivers' schedule maintenance, equipment stewardship, and cargo responsibilities respectively. Because the measures are biased potential employers cannot know drivers' employment histories and cannot evaluate the frequency of negative events or the extent of negative driver characteristics. Correspondingly these measures provide no response space or response category for drivers who perform tasks well (i.e. have only positive driver characteristics) or who experience no adverse employment events. Along with not providing for positive identifications the [redacted] measures fail to provide previous employers with categorical responses indicating whether or not drivers were culpable for negative employment events.

These comparisons establish that the [redacted] Termination Record form could have substantially more content validity by organizing its work history identifiers by categorical topics. The [redacted] Termination Record form could have substantially more measurement validity by providing comprehensive and balanced response codes. With this overall analysis in mind I now turn to an evaluation of the individual measurements in the work history section of the [redacted] Termination Record form.

Evaluation of Individual Measures. The focus of this section is the [redacted] Termination Record form work record measures, coded 901 through 999, and their respective instructions in the [redacted] Termination Record form and in the [redacted] Services Guide to the Termination Record Form. Excluding the seven codes, 950 through 956, that deal with load abandonment, there are twenty measures. Among these twenty measures 10 are unclear, 11 are inconsistent, 17 have inadequate responses, and 17 are biased. Additionally one measure, code 944 – personal contact requested, is not a measure of employment history.

In what follows I assess the content validity and measurement validity of the individual termination record form based upon categories of employment history such as safety and professional conduct. The Guide to the [redacted] Termination Record form does not provide specific guidance about these contents and therefore I created my own.

"Satisfactory, Superior, and Outstanding." These measures are intended as summary measures of employment history. [redacted] does not organize its form such that there is a

measure for “unsatisfactory employees,” but [redacted]’s instructions reflect that codes 901 through 903 are summary or aggregate identifiers.

[redacted] instructs employers to identify drivers as satisfactory if the driver “meets minimum company standards of performance in all categories”, and [redacted] instructs employers to identify a driver as superior if the driver “exceeds minimum company standards of performance in all categories.” [redacted] does not identify these constituent categories and leaves their inclusion or exclusion to carriers’ discretion. This lack of definition creates an inconsistent measure akin to including or excluding the length of children’s heads or legs for height measurements. Because each carrier may include or exclude any number of categories the definitions for satisfactory and superior are very likely to be inconsistent.

Based on [redacted]’s instructions, some companies may include only a single category (e.g. driver safety) for an evaluation of satisfactory or superior while other companies may include a dozen or more categories. Because carriers subjectively and variously define individual dimensions of drivers’ employment histories the codes *satisfactory* and *superior* lack clarity and content validity. Previous employers are free to define by inclusion and exclusion their topical components in the summary evaluations “satisfactory” and “superior” with no means of identifying those topics for potential employers. Potential employers cannot discern the components that comprise satisfactory and superior drivers. Table 2.2 explains why the codes satisfactory, superior, and outstanding have neither content nor measurement validity.

Table 2.2: Content and Measurement Validity

Work Record Item	Content Validity	Measurement Validity	Explanation
901- Satisfactory	Invalid	Invalid	<ul style="list-style-type: none"> ■ No identification of categories included (inconsistent definitions) ■ Varied thresholds for “minimum company standards” (lack of scale)
902 – Superior	Invalid	Invalid	<ul style="list-style-type: none"> ■ No identification of categories included (inconsistent definitions) ■ Varied Thresholds (lack of scale) ■ Rating may not be attainable (inadequate scale)
903 – Outstanding	Invalid	Invalid	<ul style="list-style-type: none"> ■ No identified categories (Inconsistent definition) ■ Undefined threshold for “Outstanding” (lack of scale)

Satisfactory: Driver meets minimum company standards of performance in all categories. With respect to measurement validity for code 901 satisfactory drivers must meet minimum standards but potential employers have no means of knowing what these minima are. One carrier may have a minimum standard that prohibits any swearing. A driver who said "Hell" would not be coded as satisfactory but rather as "935 company policy violation." The lack of threshold indicators will lead to inconsistent identifications.

Superior: Driver exceeds minimum company standards of performance in all categories. [redacted] instructs previous employers to identify drivers as "superior" if they "exceed" company standards. A consideration for this code is that exceeding a standard may be impossible. A company may have a standard that drivers are not to have accidents. A driver who has no accidents has met the company standard, but cannot "exceed" that standard. A driver cannot have a "negative" accident that would in some sense exceed the company standard of not having an accident. Because the measure *superior* may be unattainable it is inadequate and this decreases further its measurement validity.

Outstanding: Driver's performance is outstanding in all categories. This instruction suffers not only from the lack of categorical definition but also from a lack of reference for identifying "outstanding" which is used as both the measure and the definition of the measure. The threshold for outstanding drivers is subject to carriers' varying standards. These varied standards in turn contribute to inconsistent measurement.

Comparing the instructions for recording superior and outstanding evidences their lack of measurement validity. [redacted] instructs carriers to code as "superior" only those drivers who exceed their individual company standards for "all categories." "Outstanding" drivers are defined at the discretion of the employer – there is neither an identified threshold nor categorical definition. A carrier using a single category – safety – may code an employee as outstanding because the driver safely delivered a load through a hurricane once. Another carrier may decline to identify a driver who has served dependably for a decade but then was late for his final delivery. Because the termination record form requires that drivers exceed company standards "in all categories" and because this driver can be identified using a measure that relates to a single idiosyncratic event the previous employer may choose not to identify this superior driver as such. With clearly identified employment history topics that in turn constitute identification as satisfactory or superior potential employers could discern drivers who failed to

meet a previous employers' standards in a single category verses drivers' who failed to meet previous employers' standards in numerous categories.

Complaints: An excessive number of complaints have been received regarding the driver's service and/or safety. This measure and its instructions lack content and measurement validity. With respect to content validity, the measure "complaints" conflates reports about safety, customer service, and professional conduct. The [redacted] Termination Record form lacks guidance about relevant sources of complaints, which could include co-workers, customers, law enforcement officials, and the general public. Table 3 details the flaws in the complaint code and its instructions. The lack of a clear categorical relationship deprives the code of content validity. The lack of guidance regarding complaints about safety practices or professional conduct renders the measurement inconsistent.

The complaints code lacks a threshold for the number of complaints required to be "excessive." A potential employer can only infer that the number of complaints is more than one. This lack of an identified threshold deprives the code of measurement validity.

Table 2.3: Customer Relations Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
911 – Complaints	Invalid	Invalid	<ul style="list-style-type: none"> ■ Conflates service and safety related complaints (lack of clarity) ■ Threshold for excessive is undefined and may vary widely across carriers (inconsistent scale) ■ No categorical response available for employees without recorded complaints (Biased measure).
924 – Late Pick Up/Delivery	Valid	Invalid	<ul style="list-style-type: none"> ■ Identifies a single idiosyncratic negative event (biased measure) ■ Measure fails to identify drivers who routinely met scheduled pick ups and deliveries (Biased scale) ■ Measure makes no provision for driver's responsibility (Inadequate scale)

Late Pick Up/Delivery: Failed to make pick up or delivery according to schedule. The code Late Pick Up/Delivery could be readily identified as a constituent of customer relations or professional conduct and thus has content validity. There are three flaws that render code 924's measurement invalid. 1) Late Pick Up/Delivery identifies a single idiosyncratic negative

event. This focus biases the measure such that the driver is represented negatively no matter his or her overall record of schedule maintenance. 2) The measure makes no provision for drivers who met their pick up and delivery schedules and thus excludes most drivers. This inadequacy further detracts from the code's measurement validity. 3) The measure provides no indication of whether or not the driver was responsible for the single idiosyncratic late pick up or delivery. Some previous employers may use this code when drivers could not avoid being late (because of severe weather or unsafe conditions) while other previous employers may use this code only when lateness is clearly the driver's responsibility (e.g. the driver overslept).

Professional Conduct. The next eight codes I evaluate broadly relate to drivers' professional conduct. *[redacted]* makes no effort to identify or define this category. Table 4 presents summary evaluations for codes 913 – cargo loss, 917 – equipment loss, 915 – falsified employment application, 926 – log violation, 935 – company policy violation, 957 – unauthorized equipment use, 959 – unauthorized passenger, and 961 – unauthorized use of company funds.

Table 2.4: Professional Conduct Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
913 – Cargo Loss	Invalid	Invalid	<ul style="list-style-type: none"> ■ Could relate to customer satisfaction, safety, or employee trustworthiness (lack of clarity) ■ No assignment of responsibility for cargo lost (driver, weather/act of God, theft, equipment poorly maintained by carrier) – (Inadequate scale) ■ No response category to reflect drivers who experienced no losses (Biased measure)
917 – Equipment Loss	Invalid	Invalid	<ul style="list-style-type: none"> ■ Could relate to safety, trustworthiness, or professional conduct (Measure lacks clarity). ■ No assignment of responsibility for equipment losses (driver, weather/act of God, theft, equipment poorly maintained by carrier) – (Inadequate scale) ■ No response category to reflect drivers who experienced no losses (biased measure).
915 – Falsified Employment Application	Valid	Invalid	<ul style="list-style-type: none"> ■ Fails to record employees who submitted true employment applications (measure is biased)
926 – Log Violation	Valid	Invalid	<ul style="list-style-type: none"> ■ Fails to record employees who complied with log procedures (measure is biased)
935 – Company Policy Violation	Invalid	Invalid	<ul style="list-style-type: none"> ■ Measure fails to identify category of policies violated (e.g. customer relations, co-worker relations, professional conduct, safety) – (measure lacks clarity) ■ Measure is likely to be defined differently by carriers according to individual company policy sets (Inconsistent scale). ■ Alternative positive ratings are unavailable (measure is biased)
957, 959, 961 – Unauthorized Equipment Use/ Passenger/ Use of Company Funds	Valid	Invalid	<ul style="list-style-type: none"> ■ Measures contains no indication of culpability (Inadequate scale) ■ Measure fails to identify drivers who routinely complied with equipment, passenger, and company fund policies (Measure is biased) ■ All 3 measures rely on company standards that likely vary (scales are inconsistent/undefined).

Codes 913 Cargo Loss and 917 Equipment Loss are not clearly related to a category of employment history and thus lack content validity. These measures are inadequate because

they fail to reflect drivers' culpability for cargo and equipment losses. These measures are biased because they focus on a single idiosyncratic event and fail to record drivers who have experienced no equipment or cargo losses.

Code 915 – falsified employment application is clearly related to employee trustworthiness or professional conduct and has content validity. The measure is biased because it fails to record employees who submitted accurate employment applications.

Code 926 – log violation is related to professional conduct. Code 926 is biased because it records only drivers who have violated Federal Motor Carrier Safety Regulations, part 395. It is unclear from the instructions whether this violation relates to the driver's professional conduct, the operation of the truck, or the driver's recordation of that conduct in keeping the log. This problem will likely create inconsistent measurement.

Code 935 – company policy violation lacks content validity because the instructions state that the driver "violate(d) company policies and procedures" and that "this code to be used only when the other selections in this section do not indicate the company policy violated." Because *[redacted]* fails to identify relevant company policies and employment history topics carriers are unlikely to understand when to use this code and even less likely to discern which policies may or may not be reflected by code 935. *[redacted]*'s instructions are akin to a teacher telling a student to color a sheet of paper using any color other than the ones other students have used without permitting the student to see other students papers. The best the student can do is guess.

Codes 957 – unauthorized equipment use, 959 – unauthorized passenger, and 961 – unauthorized use of company funds are all reasonably related to professional conduct or employee trustworthiness and have content validity. The measures are inadequate because they fail to reflect culpability. The measures are biased because they fail to identify employees who complied with previous employers' policies regarding equipment use, passengers, and company funds.

Safety: Code 929 – failed to report accident and code 938 – unsatisfactory safety record relate to driver safety. Code 929 lacks direction regarding whether prospective employers should consider the code relevant to safety or professional conduct. Such direction would improve the codes content validity. Code 929 provides no response space for previous employers to indicate the context of the reporting failure and the context of the accident itself. These shortcomings render code 929 inadequate. Code 929 provides no response category

that identifies drivers who have not had accidents. Although potential employers could learn more about a driver's safety record from sections c through m of the termination record form, code 929 itself reports nothing directly about an employee's safety practices. The work record section's clarity could be improved by moving code 929 to sections c through m and by inquiry about accident reporting for each corresponding accident.

Code 929 focuses on a single idiosyncratic event, a one-time failure to report an accident, and is biased. Code 929 – failed to report accident is invalid with respect to measurement.

Table 2.5: Safety

Work Record Measure	Content Validity	Measurement Validity	Explanation
929 – Failed to report accident	Invalid	Invalid	<ul style="list-style-type: none"> ■ No instructions relating measure to safety or professional conduct (measure lacks clarity) ■ No indication of responsibility for failure to report accident (inadequate scale) ■ No response provided for drivers who report accidents according to guidelines (measure is biased) ■ No response provided for drivers who have not had accidents (measure is biased)
938 – Unsatisfactory Safety Record	Valid	Invalid	<ul style="list-style-type: none"> ■ Safety standards likely vary from company to company (Inconsistent scale). ■ Threshold for "Unsatisfactory" is undefined (unclear scale) ■ Alternative positive rating is not available (biased measure) ■ Measure focuses on a single event (biased measure).

Code 938 – Unsatisfactory Safety Record is valid with respect to its related content, driver safety. The directions state, "Driver did not meet company safety standards." The reliance on company-specific standards is likely to introduce varied scales into the measure and decrease its measurement validity. There is no threshold definition for "Unsatisfactory" other than the driver not meeting standards. The implicit threshold is a single event in which the

driver fails to meet a company safety standard regardless of the driver's culpability for that failure. The measure's inadequacy undermines its measurement validity.

There is no positive categorical response available to represent drivers who routinely met company safety standards. This bias renders the measure invalid.

The [redacted] Termination Record form contains two codes, Code 944 – personal contact requested, and Code 999 – other that have no content validity. Code 944 – personal contact requested is not a measure and thus the validity concepts do not apply. However by including code 944 along with measures [redacted] has compounded the disorganized contents of its work history section. Because many previous employers use only a single code to reflect an entire employment history this code may effectively deprive potential employers of valuable information.

Table 2.6: General Codes

Work Record Item	Content Validity	Measurement Validity	Explanation
944 – Personal Contact Requested	Not Applicable	Not Applicable	■ This code is not a measure
999 – Other	Invalid	Invalid	<ul style="list-style-type: none"> ■ Allows for measure to relate to any category of performance none of which are identified either previously or by the previous employer recording "other" (measure lacks clarity) ■ Provides no information about driver character or event histories (scale is absent)

For code 999 – other, [redacted]'s instructions are "anything other than items listed above." These instructions are so broad that the code may completely lack a relationship to employment history. The lack of categorical direction renders the measure invalid with respect to content validity. The code is not biased nor is it inadequate. The general nature of the code provides prospective employers with no indication of whether the "other" employment event or driver characteristic is positive, negative, or neutral.

In the work record section of the [redacted] Termination Record form there are nine measures (931 – Quit Under Dispatch, 933 – Quit/Dismissed During Training or Orientation, and codes 950-956) that identify single idiosyncratic events related to the context of employment termination. None of these codes provides responses reflecting drivers who did not terminate their employment under the circumstances described. The [redacted] Termination Record form

instructs previous employers to circle only one code among codes 950 through 956, but does not indicate whether doing so then limits a carrier's responsibility to use codes 901 through 944 and code 999. This confusion and lack of direction could further erode carriers' abilities to use the codes that apply to specific drivers and thus undermines the measure's scale and its measurement validity.

By way of illustration, I have produced an alternative termination record form work record section that I present as Appendix B. The ordinal categorical responses provide for greater adequacy and reduce measurement bias. If carriers used this form, they could record drivers' experiences and characteristics as they were reflected throughout employment histories. The comprehensive scales provide response categories for unsatisfactory, satisfactory, and exceptional driver characteristics, and carriers could positively identify when they have not observed a characteristic or event such that it cannot be recorded on behalf of a driver.

The twenty measures require fewer dichotomous identification choices than the current [redacted] work record form and the superior organization, with contents grouped among three topics, should facilitate accurate reporting. Given the current low use of work record codes, with 94 percent of carriers using only one work record code per termination record form, [redacted] would greatly enhance the reliability of its employment histories by creating a better organized employment history report with adequate and unbiased response scales.

Carriers could include summary measures for each topic on the alternative form. Prospective employers would have information from the individual measures within each topic such that they could evaluate whether summary topical codes of exceptional, meets expectations, or does not meet expectations comported with the individual codes recorded. [redacted] could maintain its separate section for accident reporting with additions regarding accident reporting by drivers.

In concluding this evaluation of the [redacted] Termination Record form and its instructions all of the deficiencies identified up to this point are based on the form's presentation, design, and instructions. None of the evaluations of content and measurement validity require a statistical analysis of the employment histories produced by the [redacted] Termination Record form. Rather, these design matters are typically attended *ex ante* by researchers and recorders.

Section 3: Analysis of Carrier [redacted] Form Profiles

In this section I move from a critique of the design and presentation of and instructions for the [redacted] Termination Record form to an evaluation of how consistently suppliers of statements to [redacted]'s database, previous employers, use the termination record form to reflect employment events and driver characteristics. The forthcoming analysis definitively reveals that [redacted]'s suppliers, carriers, use the termination record form codes inconsistently and the analysis strongly reflects that a majority of carriers do not have a shared understanding of when and how to use specific codes. Even a cursory non-statistical analysis identifies hundred of carriers whose code use lays beyond reasonable bounds. For example, 102 carriers have submitted over 80,000 termination record forms without once using code 901 satisfactory. Even with the articulated shortcomings in its design and the validity problems discussed in the previous section [redacted]'s managers could have and can readily assess the consistency with which motor carriers understand and employ the termination record form, and can identify any individual motor carrier or group of motor carriers who utilize termination record forms in an unusual manner. [redacted] could have and can identify carriers whose termination record forms lie beyond a reasonably expected range of code use by comparing the profiles of code usage from any single motor carrier with the profile of termination record form codes produced by a sample of one thousand or more termination record forms from the population of all [redacted]'s termination record forms.

This evaluation does not depend upon any termination record form measure being clear, unbiased, or adequate (i.e. having content and measurement validity). This evaluation assesses carriers' shared understanding of the [redacted] codes as that understanding is reflected by their code use on termination record forms they have voluntarily submitted to [redacted]. If [redacted]'s data suppliers, motor carriers, have a shared understanding of the termination record form codes and consistently use those codes to reflect driver experiences and characteristics then the profiles of the individual data suppliers' (i.e. carrier companies) termination record form code usages should not be grossly different from the code usages in the random sample of termination record forms.

We can assess the carriers' use of termination record forms by developing a range of likely percentages for carriers based upon the code usages observed in a random sample of 5000 termination record forms (hereafter the TRF sample). Statisticians refer to this range of

likely outcomes as a Confidence Interval. A confidence interval uses percentages observed in a population sample to provide a lower boundary and an upper boundary of the percentages that we would expect to observe in future or comparative samples. In this case our comparative samples are the profiles of code use for 3604 carriers that have contributed 30 or more termination record forms to [redacted].

In order to understand confidence intervals a preliminary example may be helpful. If we say that a sample of 1000 school children reports that ten percent of children age 10 are taller than 60 inches tall and that the ten percent value from the sample has a 95 percent confidence interval ranging from eight percent to twelve percent, then we are saying that in 19 out of 20 future or comparative samples of 1000 school children we expect that the proportion of children who are more than 60 inches tall will range within this confidence interval which has a lower bound of 8 percent to an upper bound of 12 percent. Notice we do not expect any single sample to estimate that exactly 10 percent of children are more than sixty inches tall. Rather we expect that for every twenty samples, 95% or 19 of them will report that the percentage of children who are more than 60 inches tall ranges from a lower bound of eight percent to an upper bound of twelve percent. The twentieth sample can then be thought of as an outlier. This means that we have no specific expectation about the percentage of children estimated to be more than sixty inches tall in that 20th sample other than that the proportion is likely to be less than 8 percent or greater than 12 percent.

If a second researcher reports ten subsequent samples of 1000 school children age 10 and four samples, 40% of the samples, estimate that more than 12% of children are taller than 60 inches, then we would estimate that three of the four samples lie beyond the expected range of percentages. If we check to ensure our initial sample is taken properly, then we would inquire about the second researcher's sampling procedure. If he reports that he sampled basketball camps then that qualification could explain why the children in his samples are reported as taller.

Before analyzing the [redacted] Termination Record form and carriers' use of individual codes within the termination record form, I develop what a confidence interval is and how it works. I first outline the components or ingredients in the confidence interval and then I walk through how it is used to evaluate reports from subsequent samples or sources.

The ingredients of a confidence interval include:

1. A population of interest
2. The property of interest from that population
3. A sampling procedure and the sample size
4. Estimated population proportions from the sample
5. Estimated standard errors for each proportion of interest
6. A confidence level

The population of interest in [redacted] is the population of termination record forms completed by carriers to represent drivers' employment histories. Notice that the population is not drivers or carriers but completed termination record forms. The TRF sample of 5000 termination record forms provides a representative sample of the population of 3.5 million termination record forms. Three thousand six hundred three carriers contributed 30 to 106,272 termination record forms to this population. On average these 3603 carriers contributed 999 termination record forms and these termination record forms comprise over 97% of the termination record forms in the population. In this analysis, we are interested in how carriers have used termination record forms to produce drivers' employment histories.

The property we are interested in from the population of termination record forms is proportion of forms that use each of the individual work record codes (i.e. codes 901 through 999). In other words what percentages of termination record forms have code 901 satisfactory marked or code 938 unsatisfactory safety record marked? We could be interested in this property for any number of reasons. For this analysis we will want to compare the percentages from the population of termination record forms, that is the profiles of code use on termination record forms, to the proportions of codes marked by individual companies.

The sampling procedure was to randomly sample 5000 termination record forms from all the 3.5 million completed forms available in the population of termination record forms. A random sample does not mean that work record forms were selected in an *ad hoc* or helter skelter fashion. A random sample means that every termination record form in the population

of termination record forms had an equal probability of being selected for the sample. If there were 500,000 termination record forms in the available population and we sampled 5000, then the probability of being sampled was 0.01. If there were 100,000 termination record forms in our sample of 5000 termination record forms then the probability of being randomly sampled was 0.05 or 5 percent. A random sample is the best sampling procedure to ensure that the sample is representative of the population (Judd, Kidder, and Smith, 1991).

Choosing a sample size is important because larger samples will provide greater precision in estimating population proportions. Larger sample sizes generate more precise estimates of population proportions (Babbie, 2004). Most survey samples of the population of the United States have a size of 900 to 1200 respondents or observations and these sample sizes result in confidence intervals that range from 3.2% below a sample proportion of 50% to 3.2% above a sample proportion of 50%. By using a sample size of 5000, we have increased our precision. In contrast to confidence intervals that are 3.2% above and below a sample proportion of 50%, our confidence intervals range from 1.4% above a sample proportion of 50% and 1.4% below a sample proportion of 50%. Notice, the sample of 5000 [redacted] Termination Record forms is four times larger than most samples of the United States population. The precision and "representativeness" of random samples does not depend on the size of the population but rather is a function of the sample size.

The fourth step in developing confidence intervals is to estimate population proportions from the TRF for each work record code. For the [redacted] work record codes the sample of 5000 produced an estimated population proportion for each code and these are reported below in table 3.1.

Table 3.1: Estimated Population Proportions from Sample of 5000 Termination Record Forms

Code	Proportion From Sample of 5000 TRF
901 Satisfactory	
912 Excessive Complaints	
924 Late Pick Up/Delivery	
926 Log Violation	
928 No Show	
931 Quit Under Dispatch	
933 Quit/Dismissed During Training/Orientation	
935 Company Policy Violation	
938 Unsatisfactory Safety Record	
944 Personal Contact Requested	
999 Other	

Table 3.1 lists the proportions we observed of code use observed from the random sample of 5000 termination record forms. For example we observed that carriers marked [REDACTED] drivers as code 901 satisfactory on the termination record form. This [REDACTED] codes used translates to an estimated population proportion of [REDACTED], and we interpret this by stating that among the three million termination record forms in the population we estimate that [REDACTED] of them are marked using code 901 satisfactory. For this analysis, I excluded those codes that were marked on less than 0.5 percent (25) of the termination record forms.

Although a sample of 5000 provides relatively precise estimates of population proportions, we now want to develop the interval in which we expect comparative sample proportions to fall. Remember we don't expect subsequent samples of 5000 termination record forms to duplicate our initial sample, but we do expect them to be within a defined range of percentages.

The next step in our procedure, noted as step 5 on page 27, is to estimate the standard error for each proportion. The standard error for sample proportions is the standard deviation of the proportion adjusted for the sample size.

The standard deviation is the average amount that sample observations differ from the sample mean. To calculate the standard deviation we follow these steps:

1. In our sample carriers marked code 901 satisfactory to denote satisfactory drivers, and these drivers were given a value of one (1) on the termination record forms in our sample. Drivers not marked as one were coded as zero (0) on code 901 on the termination record forms in our sample. The TRF sample percentage was [REDACTED] meaning that [REDACTED] termination record forms were coded as 1 on code 901 satisfactory, and [REDACTED] of termination record forms in the sample were coded as 0 on code 901 satisfactory.
2. The mean score of code 901 is equal to the percentage of termination record forms marked as code 901, which is the estimated population percentage of [REDACTED]
3. To estimate the standard deviation from the sample percentage, we multiply the TRF sample proportion, [REDACTED], by its counterpart, $1 - (\text{sample percentage}/100)$ (Diamond and Jeffries, 2001). These two terms represent the percent of drivers coded satisfactory and the percent not coded as satisfactory on our 5000 termination record forms. Then we take the square root of this product to give us the standard deviation or average distance from the estimated population percentage.
4. The formula for the standard deviation of a proportion is:

$$\text{Standard deviation} = \text{SQRT} [\text{Percentage} \times \{1 - (\text{Percentage}/100)\}]$$

Now we want to move from the standard deviation of a percentage to the standard error of a sample proportion. To do this requires that we adjust the standard deviation we just estimated to reflect the sample size. Remember, larger sample sizes provide greater precision, which means that larger sample sizes reduce the size of the standard deviation. To make this adjustment, we divide the standard deviation by the square root of the size of our sample. Our

sample is 5000 termination record forms so we will divide the standard deviation of the sample proportion by the square root of 5000. Our formula now is:

$$\text{Standard error of a proportion} = \text{SQRT} \{ \text{Percentage} \times (1 - \text{Percentage}) \} / \text{Sample Size}$$

With the standard error we can construct a confidence interval that is centered on the TRF sample proportion, which is the estimated population percentage. To continue this process we should select a Confidence Level. The confidence level provides a level of precision for our expectations about the samples or groups we wish to compare to our random and representative population sample.

A confidence level has a specific interpretation. If we choose a 95% confidence level, we are saying that we expect 19 of 20 (95%) of subsequent samples of 5000 or more observations will fall within our confidence interval. With a 90% confidence level we expect 9 of 10 samples of 5000 or more termination record forms to fall within the range of our 90% confidence interval. With a 99% confidence interval we expect 99 of 100 comparative samples to fall within the range of our confidence level. We are not saying that we are 99% sure of our estimate. We are saying that given our sample proportion, its standard deviation, and our sample size, we expect the corresponding percentage of comparative samples of that size to be within our confidence interval.

The confidence interval permits us to identify outlier samples in comparison to the population percentage estimated by the TRF sample, which is representative of the population of termination record forms. Generically, outlier samples are samples in which the reported values lie far beyond an estimated population proportion or that lie beyond the range of values from most of the samples. Statistically outlier samples can be defined as samples in which one or more reported values lay two or more standard errors outside the confidence interval. In this analysis the TRF sample generates the estimated population proportion and the termination record forms contributed by each of the 3603 carriers to the population of 3.5 million termination record forms are the subsequent comparable samples.

For every confidence level there is a value from a statistical distribution, called the Normal Distribution, that tells us the range of sample percentages that will fall above and below the estimated population percentage for which we can attribute the deviation from the population percentage to sampling variation. We can infer that sample percentages that fall

within this range are expected from sampling variation and are not caused by systemic distortions.

The range around an estimated population percentage that we must deviate to ensure that the percentage of comparative samples that fall within the confidence interval is equivalent to our confidence level. For every confidence level, we can select a corresponding "Z-score" from which we calculate the confidence interval. We can think of a Z-Score as a "distance score" that tells you how from the TRF sample percentage to estimate the boundaries of comparative samples that should fall within a range estimated by the TRF sample percentage and its standard error. Because sample sizes with 30 or more observations enable us to use the normal distribution we can use the Z-scores, or distance scores, to estimate how far from our sample proportions we must deviate to create any confidence level (Diamond and Jeffries, 2001; Smith, 1985). The equation for creating a confidence interval then is:

$$\text{Confidence Interval} = \text{Sample Percentage} \pm \{\text{Distance Score} \times \text{Sample Standard Error}\}$$

The confidence interval is centered upon the TRF sample proportion. The confidence interval has an upper bound equal to the TRF sample percentage plus the product of the distance score (Z-score) times the TRF sample percentage's standard error. The Confidence interval has a lower bound equal to the TRF sample percentage minus the product of the distance score (Z-score) times the TRF sample percentage's standard error.

The normal distribution provides consistent parameters for us to develop confidence intervals. When using the normal distribution we expect that 68.25% of all subsequent samples (i.e. those samples drawn from the same population and of the same sample size) will fall within one standard error above or below our sample percentage. Ninety-five percent of comparative samples will fall within 1.96 standard errors of our sample proportion, and 99% of comparative samples will fall within 2.575 standard errors above or below our sample proportion (Diamond and Jeffries, 2001; Babbie, 2005). Because [redacted] has not provided a its own statistical analysis of termination record form work record codes I chose the confidence level, and corresponding Z-score, that generated the widest confidence intervals and minimized the number of carriers likely, in the absence of systemic distortions, to be identified as deviants or outliers.

For example for termination record forms that use the code 901 satisfactory, we have a TRF sample percentage of [REDACTED] and a standard error of [REDACTED]. If we create an interval that is the sample proportion plus or minus 1 standard error then our interval is

$$68.25\% \text{ Confidence Interval} = [REDACTED] \pm (1) [REDACTED] = [REDACTED]\% \text{ to } [REDACTED]\%$$

If we deviate from the TRF sample percentage by plus or minus one standard error then we would expect [REDACTED]% of comparative samples to have sample percentages for code 901 satisfactory within the range of [REDACTED]% to [REDACTED]%. To create a 95% Confidence Interval, in which we expect 19 of 20 comparative samples to fall within, we have to use the distance score of [REDACTED] so that now our equation is:

$$95\% \text{ Confidence Interval} = [REDACTED]\% \pm ([REDACTED]) [REDACTED] = [REDACTED]\% \pm 1.4\% = 56.1\% \text{ to } 58.9\%$$

This 95% confidence interval has a lower bound of [REDACTED] and an upper bound of [REDACTED]%. It is wider than our [REDACTED] confidence interval. We interpret this [REDACTED] confidence interval by stating that we expect 19 of 20 subsequent samples of 5000 termination record forms will have sample proportions for code 901 satisfactory that lie between [REDACTED] and [REDACTED]%. We expect one of twenty samples of 5000 termination record forms will lie outside the interval [REDACTED]% to [REDACTED]%.

Before moving to an evaluation of the carriers' use of [REDACTED] termination record forms, two final points about confidence intervals are in order. First, higher confidence levels lead to wider confidence intervals. A 68% confidence interval ranges from one standard error below the sample proportion to one standard error above the sample percentage. A 95% confidence interval ranges from 1.96 standard errors below the sample percentage to 1.96 standard errors above the sample percentage. A 99% confidence interval ranges from 2.575 standard errors below the sample percentage to 2.575 standard errors above the sample percentage.

Second, larger sample sizes create smaller standard errors of the sample percentages and lead to smaller or tighter confidence intervals. Small sample sizes lead to larger or wider confidence intervals. A sample percentage of 40% from a sample of 1000 termination record forms has a standard error of 1.55%. A sample proportion of 40% from a sample of 30 termination record forms has a standard error of 8.94%. When estimating confidence intervals we can ensure comparability by using the smallest sample size from the comparative samples to create the confidence intervals. Larger samples will fall within the confidence intervals created

using smaller samples, but smaller samples may or may not fall within confidence intervals using larger samples.

I have used the random representative sample of 5000 termination record forms to generate the estimated population percentages. The sample of 5000 termination record forms produced the estimated population percentages for all termination record forms used to produce drivers' employment histories. The records of code use by 3603 carriers are then compared to these estimated population percentages. The 3603 carriers each has a profile that reports 1) the total number of termination record forms submitted by the carrier and 2) the frequency (i.e. the number of termination record forms) with which a carrier used each individual work record code. To obtain individual carrier code use percentages that could be compared to the TRF sample, I divided the frequency of code use by the total number of termination record forms submitted by that carrier, and then multiplied the proportion by one hundred. Each carrier's code use is then a sample that can be evaluated against the TRF sample of 5000 termination record forms and the confidence intervals generated around its percentages of code use.

Carriers submitted a minimum of 30 termination record forms and a maximum of termination record forms. The mean number of termination record forms submitted by a carrier was . The total number of termination record forms profiles in the sample of 3603 carriers is over 3.5 million.

Confidence Intervals & Evaluation of Carriers' Code Use. To evaluate how carriers use the work record codes 901 through 999 in the termination record form, I used the estimated population percentages that are presented again here as they were in table 3.1. To estimate the standard errors for these percentages, I used 30 as the sample size for all the codes. This ensures that all the confidence intervals are comparable for carriers whose profile included the minimum of 30 termination record forms. This assumption biases the evaluation in favor and means that many carriers who have used codes on an unusually small or larger percentage of termination record forms will go undetected because of the generous confidence intervals. Overall, we expect 99 of 100 the carriers in the sample of 3603 carrier profiles to fall within the confidence intervals.

Table 3.2: Termination Record Form Work Record Codes and Estimated Population Percentages

Code	Proportion From Sample of 5000 TRF	Standard Error of the Percentage (Sample size = 30 trf's or more)
901 Satisfactory		
912 Excessive Complaints		
924 Late Pick Up/Delivery		
926 Log Violation		
928 No Show		
931 Quit Under Dispatch		
933 Quit/Dismissed During Training/Orientation		
935 Company Policy Violation		
938 Unsatisfactory Safety Record		
944 Personal Contact Requested		
999 Other		

With these estimated population percentages and their standard errors for samples of size 30 or greater, I generated 99% confidence intervals for each code. These confidence intervals have Z-scores, or distance scores, of 2.575, which is from the Normal Distribution. We interpret 99% confidence intervals to mean that we expect 99 of 100 comparable samples will fall within the interval created by adding and subtracting the product of 2.575 times the standard error of the proportion to and from the sample proportion. I have converted proportions back to percentages and this conversion does not affect the results or evaluation.

After generating each confidence interval I identified how many carriers exceeded the lower or upper boundary of the confidence interval. For code 901 satisfactory I identified how many carriers' use of each code were below the lower boundary. For the remaining codes I identified how many carriers' use of each code were above the upper boundary. Among 3603 carriers, we expect [REDACTED] carriers (99% of carriers) to lie within each confidence interval for

each work record code. We expect [REDACTED] to lie either above or below the confidence interval with approximately half, [REDACTED] carriers, exceeding the interval, and approximately half, [REDACTED] carriers, lying below the interval. In cases where the interval has a lower bound below zero, I have conservatively assumed all one percent of carriers, [REDACTED], will exceed the confidence interval because they cannot fall below it.

I will first deal with code 901 satisfactory. Table 3.3 presents the estimated population percentage for code 901 satisfactory, the confidence interval for that percentage, the number of carriers whose use we expect in the absence of systemic distortions to fall below the lower bound of the confidence interval, and the number of carriers observed below the lower bound of the confidence interval.

Table 3.3: Confidence Interval for Code 901 Satisfactory

Code	Percentage From Sample of 5000 TRF	99% Confidence Interval	Number of Carriers expected to exceed Lower Bound of confidence interval	Number of Carriers Observed Exceeding Lower Bound of Confidence interval
901 Satisfactory*	[REDACTED]%	[REDACTED]	[REDACTED]	[REDACTED]

As table 3.3 reflects [REDACTED] of termination record forms in the sample of 5000 termination record forms were coded as satisfactory. Subtracting and adding the product of Z-score of 2.575 times the standard error, [REDACTED], from the estimated population percentage, [REDACTED]% generated the confidence interval for this code. The lower bound of this confidence interval is [REDACTED]%. The upper bound of this confidence interval is [REDACTED]%. We expect [REDACTED] of [REDACTED] carriers to have identified less than [REDACTED] of the drivers coded on their termination record forms as satisfactory by using code 901. In the 3603 carrier samples we observe that [REDACTED] carriers use code 901 satisfactory on then [REDACTED]% of their termination record forms.

To evaluate this result recall that we expect one-half of one percent of carriers to have profiles falling below the confidence interval. We observe [REDACTED] of 3603, or [REDACTED], carriers using code 901 satisfactory at rates that fall below the lower bound of the confidence interval. In other words *there are 33 times as many carriers below the confidence interval's lower bound as we expected* given our estimated population percentage, our 99% confidence level (which

widens the confidence interval to bias the interpretation in [REDACTED] favor) and our assumption that every carrier contributed only 30 termination record forms (which further widens the confidence interval to bias the analysis in [REDACTED] favor). From the result that [REDACTED] carriers lay below the confidence interval we can infer that [REDACTED] % of carriers contributing termination record forms to [REDACTED] in ways that are systematically different from ways the estimated population percentages for [REDACTED] termination record forms. Although a specific cause cannot be isolated, there are systemic distortions in carriers' use of code 901 satisfactory.

These statistically significant deviations, occurring among one-sixth of carriers in the carrier sample, are based upon comparing individual carriers' use of [REDACTED] termination record forms with the TRF sample. The producers of both samples are carriers who have been drivers' previous employers. We are not comparing carriers' use of [REDACTED] termination record forms to drivers or to driver's self-reports about employment history. In other words, this analysis compares apples to apples or carrier produced termination record forms to carrier-produced termination record forms.

The frequency with which carriers deviated from the estimated population percentage, with one in six carriers falling below the lower bound of the confidence interval should have been plainly evident to [REDACTED]. In the sample of [REDACTED] carriers ([REDACTED] carriers) used code 901 satisfactory zero (0%) of the time! Although these carriers represent slightly less than three percent of all carriers they represent nearly *six times* the expected frequency of carriers using code 901 satisfactory on zero to thirty-four percent of their termination record forms. These [REDACTED] carriers submitted between [REDACTED] and [REDACTED] termination record forms. Collectively these [REDACTED] carriers submitted [REDACTED] termination record forms never using code 901 satisfactory. Any casual observer, including [REDACTED] managers, could identify these carriers without performing a statistical analysis.

The decision to use a 99% confidence interval and to assume only 30 termination record forms per carrier for purposes of estimating the confidence intervals means that we are very likely underestimating the number of carriers who fall below the confidence interval. For subsequent evaluations of individual codes, I will be underestimating carriers' code usages that fall above the upper bound of the 99% confidence intervals.

Table 3.4 presents the estimated population percentage for work record codes 912 through 999, the confidence interval for each percentage, the number of carriers we expect to

exceed the upper bound of the confidence interval and the number of carriers observed to exceed the upper bound of the confidence interval.

Table 3.4: 99% Confidence Interval for [REDACTED] Work Record Codes 900 through 999

Code	Percentage From Sample of 5000 TRF	99% Confidence Interval	Number of Carriers expected to exceed Upper Bound of confidence interval	Number of Carriers Observed Exceeding Upper Bound of Confidence interval
912 Excessive Complaints	[REDACTED]%	0% [REDACTED]	[REDACTED]	[REDACTED]
924 Late Pick Up/Delivery	[REDACTED]%	[REDACTED]%	[REDACTED]	[REDACTED]
926 Log Violation	[REDACTED]%	[REDACTED]	[REDACTED]	[REDACTED]
928 No Show	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
931 Quit Under Dispatch	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
933 Quit/Dismissed During Training/Orientation	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]
935 Company Policy Violation	[REDACTED]%	[REDACTED]	[REDACTED]	[REDACTED]
938 Unsatisfactory Safety Record	[REDACTED]%	[REDACTED]	[REDACTED]	[REDACTED]
944 Personal Contact Requested	[REDACTED]%	[REDACTED]	[REDACTED]	[REDACTED]
999 Other	[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED]

For the remaining work record codes we expect no more than [REDACTED] carriers to have profiles in which they use an individual code on a greater percentage of termination record forms than has been indicated by the upper boundary of the 99% confidence interval. We observe the following:

1. For code 912 Excessive Complaints, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]%. This is 8.7 times greater than expected.

2. For code 924 Late Pick Up/Delivery, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **16.2** times greater than expected.
3. For code 926 Log Violation, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]%. This is **14.2** times greater than expected.
4. For code 928 No Show, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]%. This is **9.4** times greater than expected.
5. For code 931 Quit Under Dispatch, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]%. This is **12.9** times greater than expected.
6. For code 933 Quit/Dismissed During Training, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **5.6** times greater than expected.
7. For code 935 Company Policy Violation, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **8.6** times greater than expected.
8. For code 938 Unsatisfactory Safety Record, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **4.0** times greater than expected.
9. For code 944 Personal Contact Requested, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **16.1** times greater than expected.
10. For code 999 other, [REDACTED] carriers are above the upper boundary of the confidence interval of [REDACTED]. This is **6.4** times greater than expected.

For all ten of these work record codes and code 901 satisfactory the carrier profiles from the sample of 3603 carriers unmistakably reveal that substantial numbers of carriers systematically use work record codes in ways that dramatically deviate from reasonable statistical expectations. This analysis does not provide the means to impute a cause for these statistically significant deviations and systemic distortions from the expected range, but I cannot stress enough that these deviations are very substantial and very frequent. This analysis very strongly reflect that [REDACTED] suppliers, drivers' previous employers, share little if any understanding of the [REDACTED] termination record form codes and under what circumstances to use which codes.

Even after making every effort to estimate the widest possible confidence intervals we observe a minimum of [REDACTED] of [REDACTED] carriers using codes beyond the expectation. Not a single code of the eleven codes evaluated has carrier profile percentages that fall within the confidence intervals estimated with the sample of 5000 termination record forms.

One response to this analysis could be that a minority of carriers, perhaps 15% of carriers, could be creating these frequent and substantial deviations by always using codes based on individual deviant practices.² This is not the situation. Using the SPSS software I identified carriers that fell outside of the confidence interval for one or more of the codes in tables 3.3 and 3.4. Among [REDACTED] carriers [REDACTED] fell outside of one or more confidence intervals leaving only [REDACTED] that were consistently within the confidence intervals. In other words [REDACTED] of carriers in the carrier sample demonstrated deviant coding for at least one work record code. This assessment understates carriers' coding problems because it is based on the widest possible confidence intervals. A carrier-by-carrier analysis would reveal that an even larger percentage of carriers fall beyond estimated confidence intervals because such an analysis would adjust for the number of termination record forms submitted by each carrier in developing the confidence intervals.

If I were a business manager identifying sources of information that lay beyond statistically reasonable expectations I would recommend using a tighter 95% confidence interval that would provide an expectation of 19 of 20 carriers coding within the confidence interval and I would adjust the confidence intervals for the number of termination record forms submitted by individual carriers rather than using a minimum of 30 which results in generous confidence intervals. These two adjustments would reveal that more than [REDACTED] carriers used one or more work record codes with frequencies that exceeded the bounds of estimated confidence intervals.

This analysis reflects the lack of content and measurement validity of carriers' use of the termination record form and employment histories. With respect to content validity carriers' wide ranging use of codes reflects a lack of shared understanding that goes beyond what we would expect from the variable experiences drivers are likely to have. With respect to measurement validity the most problematic information comes from the large number of carriers who infrequently code drivers as "satisfactory" and the large number of carriers who

² "Deviant" here refers to statistical departures from the expected range and is not meant to provide a normative connotation.

frequently use codes such as 926 log violation, 933 quit/dismissed during training or orientation, and 938 unsatisfactory safety record. If these measures were valid and representative of drivers' employment histories one would have to inquire how companies could remain viable with less than one-third of their employees being satisfactory and in observations carriers have operated with no satisfactory drivers.

Section 4:

Reliability in the [redacted] Form Work Record Section

The [redacted] Termination Record form's "Work Record" section has twenty-seven items that together present employment histories of drivers employed by carriers who contribute termination record forms to [redacted]. An important consideration for [redacted] and potential employers is whether the employment histories presented are **reliable** representations of drivers' employment histories. Reliability is the extent to which a survey, questionnaire, test, or other measurement instrument *consistently* depicts the same events and characteristics (Carmines and Zeller, 1979; Lewis, 1999; Litwin, 1995; Thompson, 2003). A lay definition of reliability is the degree to which repeated measurements, or measurements taken under identical circumstances, will yield the same results. A statistical definition of reliability is that reliability is a measure of the randomness of the measurement process itself. Reliable measures minimize the random or variable component of the measurement process (Lewis, 1999).

An example may bring the concept of reliability into stark relief. Suppose I weigh myself on my bathroom scale one morning and the scale reads 164 pounds. Ten minutes later I step on the scale again and it reads 165 pounds. Is this scale unreliable? Probably not. The one-pound difference may be attributable to a number of factors – perhaps I put my shoes on between the measures, perhaps I ate a large breakfast, or perhaps the reading changed because of how I centered myself on the scale.

Next assume I weigh myself on a second scale the next day and it reads 163 pounds. Ten minutes later I check again and the scale reads 191 pounds! To make sure I didn't rig the scale or misread it, I step on the scale a third time. Now the scale reads 147 pounds! Three readings with a range of 44 pounds over ten minutes cannot be attributed to any of the factors that may have caused the one pound difference the previous day on the first scale. I would infer that a scale that provides three readings with a range of 44 pounds is unreliable. Because this bathroom scale is unreliable it produces inaccurate measures of my weight. Unreliable surveys and questionnaires produce inaccurate information and unreliable employee evaluation forms produce inaccurate employment histories (Brown, 2005; Litwin, 1995).

Before moving to an analysis of the [redacted] Termination Record form's reliability I note the frequency with which carriers fail to use work record codes at all. Before we can rely

termination record form. Termination record forms that use none of the work record codes are incomplete and inaccurate. Carriers who do not use the work record codes mean cannot produce work histories. Among 5000 termination record forms in a sample of termination record forms submitted after July 1, 1999, carriers used no work record codes (number 901 through 999) for termination record forms. In other words of termination record forms failed to produce work histories. Even if all twenty-six work record codes number 901 through 961 are inapplicable to a driver's employment history, the carrier has the opportunity to so indicate by using the code 999 – other in the work record section. This initial error rate could have been and can be readily observed by

instructs carriers to use all codes that apply to drivers' employment histories. Nevertheless the vast majority of carriers use only one code per driver. This has the effect of decreasing the variance of the individual codes and substantially decreasing the overall summed "score" of experiences and characteristics for each driver. Despite's claims that it views the work record codes as "issue identifiers" its instructions to check all codes that apply are instructions to produce complete employment histories. In the sample of 5000 termination record forms, only termination record forms, had two or more 900 codes checked. Among drivers not coded as satisfactory, superior, or outstanding, only out of had two or more work record codes identified. Among these same drivers drivers had no 900 codes used and no employment history produced on their behalf by the work record section of the termination record form. Carriers are more likely to represent that drivers had no work experiences, including "other," than they are to identify drivers as having one or more work experiences or driver characteristics. The inference from this data summary is that most termination record forms produce incomplete, and therefore inaccurate, employment histories.

In addition to the termination record forms with no 900 codes used, there were termination record forms on which employers used only code 944, personal contact requested, or codes 950 through 956, which identify the location and circumstance of termination. These eight codes, taken alone or together, do not provide information that can then be evaluated as an employment history. Code 944 is not a measure but a request. Codes 950 through 956 deal with a final idiosyncratic event during employment but are not sufficiently informative to denote an employment history. Of 5000 termination record form carriers used too few codes for potential employers to denote an employment history.

For a more complete analysis of the termination record form's reliability, I will use a procedure developed by Frederic Kuder and M. Richardson in 1937 (Kuder and Richardson, 1937). This procedure estimates reliability for surveys and forms in which measures are dichotomously identified. Dichotomous measures include categorical responses such as yes/no, agree/disagree, and applies/does not apply. In the sample of 5000 termination record forms provided by [redacted], identified codes were denoted as one (1) and those left blank or not used were denoted as zero (0). The use of the Kuder-Richardson 20 procedure and formula is appropriate for estimating the [redacted] Termination Record form's reliability as it relates to drivers' employment histories (Litwin, 1995; Thompson, 2003). The Kuder-Richardson 20 procedure estimates the consistency with which carriers use the work record section codes to produce employment histories.

The Kuder-Richardson 20 procedure includes a formula, commonly called the KR-20, that estimates data reliability. The product of the KR-20 procedure, which is called the KR-20 coefficient or KR-20 estimate of reliability, ranges from zero to one as an indicator of reliability.³ KR-20 estimates equal to zero indicate completely unreliable data and KR-20 estimates equal to 1 indicate completely reliable data (Brown, 2005; University of Connecticut, 2005). An example demonstrates how the KR-20 formula is used to calculate the KR-20 statistic.

We begin with ten employees and the history of their employment training. There are six experiences that the employer records. Each employee is scored as a one if he or she has had the experience and zero otherwise. Thus we might have a table of employees and their experience codes that looks like this:

³ The Kuder-Richardson 20 formula and its resulting reliability coefficient are very closely related to the reliability formula developed by Cronbach and its resulting reliability coefficient, which is commonly referred to as Cronbach's Alpha (Thompson, 2003). KR-20 is actually actually a special case of Cronbach's Alpha for data that are dichotomously coded (Carmines and Zeller, 1979; Thompson, 2003).

Table 4.1: Employee Experience Codes

Employee	1	2	3	4	5	6
Alice	0	0	0	0	0	0
Bob	0	0	0	0	1	0
Carol	1	0	1	1	1	0
David	1	1	1	1	1	1
Edith	1	1	1	1	1	1
Fred	0	0	1	0	0	0
Gail	0	0	1	1	1	0
Henry	0	0	0	1	0	0
Irene	1	0	1	1	1	0
John	0	1	0	1	0	1

Next we sum the codes for each employee going across the row such that there scores can range from 0 (no experiences recorded) to 6 (all experiences recorded). We now have the following scores for each employee:

Table 4.2: Employee Experience Codes & Summed Scores

Employee	1	2	3	4	5	6	Total
Alice	0	0	0	0	0	0	0
Bob	0	0	0	0	1	0	1
Carol	1	0	1	1	1	0	4
David	1	1	1	1	1	1	6
Edith	1	1	1	1	1	1	6
Fred	0	0	1	0	0	0	1
Gail	0	0	1	1	1	0	3
Henry	0	0	0	1	0	0	1
Irene	1	0	1	1	1	0	4
John	0	1	0	1	0	1	3

The employees' scores range from Alice's 0 to David and Edith's 6's. The mean employee score is the sum of the scores, 29, divided by the number of employees, 10. The mean experience score is 2.9. We now subtract the mean score from each employee's individual score. The resulting differences are each employee's deviation from the mean.

Table 4.3: Employee Scores & Deviation from the Mean

Employee	Total	Mean	Deviation
Alice	0	2.9	-2.9
Bob	1	2.9	-1.9
Carol	4	2.9	1.1
David	6	2.9	3.1
Edith	6	2.9	3.1
Fred	1	2.9	-1.9
Gail	3	2.9	0.1
Henry	1	2.9	-1.9
Irene	4	2.9	-1.1
John	3	2.9	0.1

We now want to square each deviation such that the negatives become positive. If we leave the deviations as they are they will simply cancel one another out.

Table 4.4: Employee Scores, Deviation & Squared Deviation

Employee	Total	Mean	Deviation	Squared deviation
Alice	0	2.9	-2.9	8.41
Bob	1	2.9	-1.9	3.61
Carol	4	2.9	1.1	1.21
David	6	2.9	3.1	9.61
Edith	6	2.9	3.1	9.61
Fred	1	2.9	-1.9	3.61
Gail	3	2.9	0.1	0.01
Henry	1	2.9	-1.9	3.61
Irene	4	2.9	-1.1	1.21
John	3	2.9	0.1	0.01

We now sum the column of squared deviations, which totals 40.9. This total is called the "Sum of Squares." We divide this number by the number of employees minus 1 ($10-1=9$). The resulting quotient of the Sum of Squares divided by the number of employees less one is 4.544. This quotient is called the variance of the experience scores.

Next we calculate the proportion of employees who the employer recorded as having each individual experience. This proportion is called p . We next will calculate the proportion of employees not having an experience. This is simply $1 - p$ but can be verified by counting the number of zeroes in each column and dividing this count by 10, the number of employees.

Table 4.5: Employee Experience Codes, Scores & Proportion Having/Not Having an Experience

Employee	1	2	3	4	5	6	Total
Alice	0	0	0	0	0	0	0
Bob	0	0	0	0	1	0	1
Carol	1	0	1	1	1	0	4
David	1	1	1	1	1	1	6
Edith	1	1	1	1	1	1	6
Fred	0	0	1	0	0	0	1
Gail	0	0	1	1	1	0	3
Henry	0	0	0	1	0	0	1
Irene	1	0	1	1	1	0	4
John	0	1	0	1	0	1	3
Proportion having experience	0.4	0.3	0.6	0.7	0.6	0.3	
Proportion not having experience	0.6	0.7	0.4	0.3	0.4	0.7	

We next calculate the product of the proportion of employees having an experience recorded with the proportion not having the experience recorded. These products are:

Table 4.6: Products of Proportions

Task	Proportion Having Experience	Proportion Not Having Experience	Product
1	0.4	0.6	0.24
2	0.3	0.7	0.21
3	0.6	0.4	0.24
4	0.7	0.3	0.21
5	0.6	0.4	0.24
6	0.3	0.7	0.21

We can now move to the KR-20 formula which is as follows:

$$KR-20 = [\#of\ Codes / (\#of\ Codes - 1)] \times [1 - (Sum\ of\ "proportion\ products" / Variance)]$$

For our example the number of codes is the number of experiences, or 6. The “Proportion Products” are above and sum to 1.35 (0.24+0.21+0.24+0.21+0.24+0.21). The variance is from table 4.4 and is 4.544. Plugging these numbers into the KR-20 formula yields the following:

$$KR-20 = (6/5) \times [1 - (1.35/4.544)] = 1.2 \times [1 - 0.297] = 1.2 \times 0.703 = 0.84$$

Our KR-20 estimate of the reliability of the experiences coded by the employer is 0.84. Now we may ask whether this is good or bad. KR-20 ranges from 0 to 1 and increasingly positive estimates reflect increasing reliability (Brown, 2005). Generally, social scientists consider surveys and records with reliability scores above 0.7 to be fairly reliable and those above 0.8 to be reliable (Braumoeller and Goertz, 2000; Brown, 2005; Carmines and Zeller, 1979; Litwin, 1995). A second way of evaluating the KR-20 estimate is to estimate an *error rate* (Braumoeller and Goertz, 2000). This is done straightforwardly and is defined as:

$$1 - SQRT(KR-20) = 1 - SQRT(0.84) = 1 - 0.916 = 0.084$$

With the records of employee experiences we have an estimated reliability of KR-20 which equals 0.84 and we have an estimated error rate of 8.4%. We would assess this coding as reliable.

Turning now to the [redacted] Termination Record form we can follow the same procedure to estimate its reliability. Now our sample provides 5000 termination record forms submitted by drivers’ previous employers. The characteristics and experiences that are included in the KR-20 estimate include the following:

**Table 4.7: Termination Record Form Codes
& Frequency of Code Use (from sample of 5000)**

Code	Frequency of Use	Percent of 5000 TRF's
912- Excessive Complaints		
913 – Cargo Loss		
915 – Falsified Employment Application		
917 – Equipment Loss		
924 – Late Pick Up/Delivery		
926 – Log Violation		
928 – No Show		
929 – Failed to Report Accident		
931 – Quit Under Dispatch		
933 – Quit/Dismissed During Training/Orientation		
935 – Company Policy Violation		
938 – Unsatisfactory Safety Record		
957 – Unauthorized Equipment Use		
959 – Unauthorized Passenger		
961 – Unauthorized Use of Company Funds		
999 – Other		

In reviewing the termination record form's work record section, one may note that other codes are included on the form but not in my analysis. This is for two reasons. First including the codes 901 – satisfactory, 902 – superior, and 903 – outstanding, produces negative KR-20 coefficients.⁴ Negative coefficients were also estimated when codes 950 through 956 and code

⁴ The interitem correlations among codes 901 through 903 and codes 911 through 999 were negative. Among 51 inter-item correlations, 50 were negative. The only exception was the correlation among code 903 – Outstanding and code 933 – Quit/Dismissed during Training or Orientation. In estimating either KR20 or Cronbach Alpha coefficients one expect positive inter-item correlations. These negative inter-item correlations reinforce the decision to use KR-20 as it requires the least transformation of data.

944 were included. KR-20 ranges from zero to one as an indicator of reliability, and negative products of the formula indicate substantial, systematic problems with the data but cannot be considered estimates of reliability.⁵ Therefore codes 901, 902, 903, code 944, and codes 950 through 957 are excluded from the analysis.

A second reason to exclude codes 950 through 956 and code 944 is that these codes do not represent the presence of employee's character or their experience but rather provide limited information about drivers' termination circumstances.

The work record descriptors included in the analysis include the 17 listed in table 7. For every termination record form I added the number of work record descriptors used and this then became the work experience score for each termination record form. This score ranges from zero to seventeen. Despite having seventeen work record codes available, no carrier used more than eight scores on an individual termination record form. Table 8 indicates the frequency of termination record form work record usage.

Table 4.8: TRF Work Record Code Scores

Score	Frequency	Percent
0		
1		
2		
3		
4		
5		
6		
7		
8		

Using the values from the above tables, we can now estimate the KR-20 reliability coefficient. The KR-20 formula is:

$$KR-20 = [\# \text{ of Codes} / (\# \text{ of Codes} - 1)] \times [1 - (\text{Sum of "proportion products"} / \text{Variance})]$$

The values from the tables can be inserted such that the formula becomes:

$$KR-20 = (\text{ }) \times [1 - (\text{ })] = \text{ } \times [1 - \text{ }] = \text{ }$$

⁵ See Thompson (2003) pp. 12-14.

The KR-20 estimate for the reliability of the work record descriptors in the termination record form is 0 well below accepted levels of 0.7 or 0.8 (Brown, 2005; Carmines and Zeller, 1979; Litwin, 1995). The estimated error rate that results from this KR-20 formula is $1 - \text{SQRT}(0.15)$ which is 0.39 or 39%! Thus the data representing employment histories on termination record forms submitted by carriers is very, very unreliable.⁶

The basic analysis of carriers' use of work record codes and the statistical analysis using Kuder-Richardson 20 reveal that the termination record form work record section produces very unreliable data and very unreliable employment histories. Because the data are very unreliable they cannot be considered valid in their current form. Data reliability is a necessary but not sufficient condition for validity. This finding does not detract from the analysis detailing the content invalidity and measurement invalidity of the termination record form work record section. This finding reinforces the points made about invalidity in the previous sections.

An alternative interpretation of the KR-20 coefficient of 0.15 focuses on statistical "unreliability." Unreliability coefficients are 1 minus any the reliability coefficient, in this analysis KR-20 (Brown, 2005; Lewis, 1999). In this case we would interpret 0.15 as indicating that the data produced by the termination record form are 85 percent unreliable. The more conservative interpretation is to use the error rate of 60 percent (Braumoeller and Goertz, 2000).

⁶ In a separate analysis, I transposed codes 901 to 903 and codes 911 through 999 into a "satisfaction index." I transposed the codes relating to rehire eligibility into a "rehire index." The Cronbach's Alpha for the two indexes was 0.541. However when the rehire index is compared to the individual identifiers, which claims are comprehensive, Cronbach's Alpha falls to 0.28. One contributing factor to this decrease in the estimated reliability is that 94% of carriers use only one or zero work record codes. If termination record forms are incomplete then correlative relationships among the work record codes are suppressed which decreases the data reliability regardless of whether it is evaluated using KR-20 or Cronbach's Alpha.

Section 5: Conclusion

The preceding four sections establish that the [redacted] Termination Record form work record section lacks content validity and measurement validity and that the data produced by [redacted] represented as employment histories are unreliable. The statistical analyses presented in sections 3 and 4 are standard and readily available using either SPSS or Stata software and could be conducted using a Microsoft Excel spreadsheet. Even in the absence of these straightforward statistical analyses, [redacted] managers could have and should have readily observed anomalies that have been evident in their data since 1999. In this conclusion, I present specific findings from sections 2, 3, and 4.

Section 2 established that the [redacted] Termination Record form lacks content validity and its individual measures are invalid because their scales are inadequate and their response categories are biased. Among twenty work record descriptors, 10 are unclear and lack content validity. Eleven of twenty measures have inconsistent thresholds for identification, seventeen lack adequate response categories, and seventeen are biased. [redacted] can easily repair much of this deficiency without adding any additional work for carriers by providing unbiased and adequate response categories and clear directions identifying thresholds for measurement categories.

The lack of clarity in [redacted]'s work "identifiers" creates confusion about the topics to which the measures relate. [redacted] could group identifiers, or measures, according to three to six clearly identified topics and this identification would likely facilitate easier use of the termination record form along with more consistent use of the individual codes. This confusion, which is compounded by the disorganized ad-hoc code presentation in the work record section, renders the work record section invalid with respect to its contents.

The extremely high prevalence of biased and inadequate measures, with 17 of 20 measures being biased and inadequate, renders the individual measures invalid. Because so many identifiers are biased and inadequate and because [redacted] instructs carriers to use all codes that apply there can be no doubt that the employment histories [redacted] produces from termination record forms are biased and inadequate. These employment histories are inaccurate and cannot be compared to one another. There is no means of assessing the relative frequency of events or the relative extent of driver characteristics.

2 The analysis in Section 3 reflects the lack of shared understanding among carriers about how and when to use work record codes. Unclear topical relationships among codes (i.e. content invalidity), biased and inadequate individual measures (i.e. invalid measurements), and incomplete and miscoded work record sections (i.e. unreliable data) all contribute to this lack of understanding. Carriers' misuse of work record codes can and should be easily observed and inferred. With over 100 carriers never using code 901 satisfactory and with hundreds of carriers using codes with frequencies that exceed expectations based on sampling and reflect systemic distortions, [REDACTED]'s managers could have readily observed anomalies in its employment histories and they could have addressed these anomalies by changing the termination record form's measures and refining its directions.

2 The analysis in Section 4 reflects the lack of reliability in the employment histories [REDACTED] produces using the termination record form work record section. The KR-20 coefficient for the work record descriptors is a paltry [REDACTED], well below any reasonable expectation above 0.7 or even 0.5. The estimated error rate of 60% is high and suggests that the work record measures are grossly unreliable. This finding is reinforced by the extraordinary frequency with which carriers use only a single work record code including codes relating to load abandonment. Because [REDACTED] percent of termination record forms have only one or zero work record codes used we can infer that only [REDACTED] percent of drivers' employment histories contain multiple recordings of driver characteristics and experiences.

As a result of the lack of content validity, the invalid measurements, and the unreliable data produced by termination record forms drivers' employment histories are inaccurate and fail to reflect their range of experiences, the extent of characteristics, and their overall employment experience. Given the multiple and compounding flaws in the [REDACTED] termination record form work record section, I would assess that employment histories are much more likely to be incomplete and therefore inaccurate than they are likely to be complete and accurate. The [REDACTED] termination record form work record section is very likely to produce inaccurate employment histories and I can envision no circumstance under which it would provide the maximum possible accuracy.

3 Finally the termination record forms evaluated in this report were submitted beginning in 1999 through the summer of 2004. The consequences of the termination record form's validity and reliability shortcomings have been evident for five years or longer. The patterns of code misuse and inaccurate or incomplete identification are ubiquitous across measures and persist

over time. Even modest changes in the measures scales and presentation format would yield substantial improvements in accuracy, validity, and reliability.

References

- Babbie, Earl. 2005. *The Basics of Social Research, 3rd Edition*. Belmont, CA: Thomson-Wadworth.
- Babbie, Earl. 2004. *The Practice of Social Research, 10th Edition*. Belmont, CA: Thomson-Wadworth.
- Braumoeller, Bear F. and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44(4):844-858.
- Brown, James Dean. 2005. *Statistics Corner: Reliability of Surveys*. University of Hawaii.
- Diamond, Ian, and Julie Jeffries. 2001. *Beginning Statistics: An Introduction for Social Scientists*. Thousand Oaks, CA: Sage Publications.
- Fowler, Floyd J., Jr. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.
- Judd, Charles M., Eliot R. Smith, and Louise H. Kidder. 1991. *Research Methods in the Social Sciences, 6th Edition*. Chicago, IL: Holt, Rinehart, and Winston, Inc.
- Lewis, Roger J. "Reliability and Validity: Meaning and Measurement." Presented at the 1999 Annual Meeting of the Society for Academic Emergency Medicine. Boston, MA.
- Litwin, Mark S. 1995. *How to Measure Survey Reliability and Validity*. Thousand Oaks, CA: Sage Publications.
- Kuder, G.F. and M. W. Richardson. 1937. The Theory and Estimation of Test Reliability. *Psychometrika*, 2:151-160.
- Smith, Gary. 1985. *Statistical Reasoning*. Boston, MA: Allyn & Bacon, Inc.
- Thompson, Bruce. 2003. "Understanding Coefficient Alpha, Really." In *Score Reliability: Contemporary Thinking on Reliability Issues*, Bruce Thompson, ed. Thousand Oaks, CA: Sage Publications.
- University of Connecticut. Kuder-Richardson 20 (KR-20) Formula. Web-based presentation available at www.gifted.uconn.edu/siegle/instrument_reliability_and_validity.htm
- United States Department of Labor. "Dictionary of Occupational Titles, Definitions for 904.383-101 Tractor Trailer-Truck Driver and 905.663-014 Truck Driver." Washington, DC.

Appendix A
Employee Evaluation Forms for
Rockingham County (VA) Public Schools,
Tacoma Public Schools, and the
***[redacted]* Form**

Rockingham County Public Schools
Staff Performance Evaluation Instrument
Bus Driver

Employee's Name: _____

School: Choose one:

Name of Evaluator: _____

Date: _____

I. Organization and Planning

<u>Evaluation of Performance</u>	<u>Outstanding</u>	<u>Satisfactory</u>	<u>Needs Improvement</u>	<u>Does Not Meet</u>
A. Is punctual and regular in attendance	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Utilizes work time efficiently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. Performs all assigned duties promptly, accurately and effectively	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Demonstrates initiative and good judgment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Performs other duties as appropriate and/or directed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments:

II. Interpersonal Relations and Communication

<u>Evaluation of Performance</u>	<u>Outstanding</u>	<u>Satisfactory</u>	<u>Needs Improvement</u>	<u>Does Not Meet</u>
A. Creates an inviting and professional atmosphere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Maintains a cordial and effective relationship in meeting the public	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. Demonstrates courtesy and professionalism in all communications	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Appropriately and effectively communicates with co-workers, teachers, administration, students, parents, and the general public	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments:

III. Professional Responsibilities/Qualities

<u>Evaluation of Performance</u>	<u>Outstanding</u>	<u>Satisfactory</u>	<u>Needs Improvement</u>	<u>Does Not Meet</u>
A. Maintains effective working relationships with other employees; works effectively as a team member; maintains positive attitude.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Upholds standards of confidentiality	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
C. Demonstrates willingness to pursue professional development, training and growth opportunities	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Observes and promotes safe work practices	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Demonstrates flexibility in work assignments and schedule; is available for altered work schedules.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F. Demonstrates problem-solving skills and abilities	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
G. Adheres to and promotes RCPS School Board policies and procedures	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Comments:

IV. Knowledge and performance of job responsibilities

<u>Evaluation of Performance</u>	<u>Outstanding</u>	<u>Satisfactory</u>	<u>Needs Improvement</u>	<u>Does Not Meet</u>
A. Maintains a consistent time of arrival at and departure from school.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Handles discipline problems effectively.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. Practices emergency evacuation drills according to policy.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
D. Transports only students assigned to bus unless otherwise authorized by principal with consent of parents.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Adheres to all state and local laws and policies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F. Performs pre- and post-trip inspections thoroughly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

G. Demonstrates safe and defensive driving skills.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
H. Maintains bus in a clean condition.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
I. Submits accurate, neat reports promptly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J. Operates school bus in accordance with route schedule and designated stops.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
K. Keeps a copy of computerized bus roster in bus and adheres to.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
L. Reports accidents, delays and driver absences to Transportation Office promptly.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
M. Reports violators of school bus stop law promptly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N. Reports mechanical problems to bus garage foreman promptly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Comments:

I have read this evaluation.

Evaluator's Signature

Evaluatee's Signature

BUS DRIVER EVALUATION**Tacoma**
Public Schools

	Exceptional	Meets	Exceeds	Needs	Improvement	Not Observed
<i>Job Knowledge/Skills</i>						
Performs pre-trip/post-trip inspection of bus to ensure it is in safe operating condition						
Keeps interior/exterior of bus clean						
Demonstrates good technical driving skills						
Is courteous, careful, attentive and alert while driving						
Skill in driving in all weather and road conditions						
Maintains appropriate schedule(s) for route(s)						
Maintains a clear and thorough route book						
Skilled in following directions and maps						
Completes Vehicle Condition Reports (VCR) as needed						
Complies with state, local and district regulations						
Assists physically handicapped children in boarding and departing from the bus as needed						
Demonstrates safe lifting procedures						
<i>Student Management Skills</i>						
Disciplines students using positive reinforcement						
Is cooperative, considerate, tactful and sensitive in managing students						
Deals with emergencies effectively						
Communicates effectively with teachers on discipline of students						
<i>Personal Qualities</i>						
Recognizes individual differences when disciplining students						
Does not lose temper or argue with students, parents, staff or co-workers						
Maintains temper and professionalism when dealing with citizens						
Is courteous to others						
Communicates positively while using the 2-way radio						
Cooperates with building administrators & transportation staff, keeping them informed of needed changes						
Adjusts easily to new assignments or changing working/driving conditions						
Reports to work on time						
Attendance						
Dress appropriately—neat, clean and well-groomed						

Name _____ Period of report _____

Date _____ Job title _____

1 - Evaluator's comments:

1. Employee's comments:

Evaluator's signature _____

Date _____

The signature below does not necessarily imply that the employee agrees with the preceding report, but only that he/she has seen and discussed it with the evaluator and/or supervisor

Employee's signature _____

Date _____

TERMINATION RECORD

Credit Value \$2.40

1 MEMBER I.D. #

2 INDIVIDUAL'S LAST NAME _____ FIRST NAME INITIAL _____

3 SOCIAL SECURITY # _____ 4 PERIOD OF SERVICE: from _____ to _____ 5 # OF ACCIDENTS _____

6 STATE _____ LICENSE # _____

STATE _____ LICENSE # _____

IMPORTANT NOTICE: Refer to "Guide" for full explanation of codes below (Form G0993)

7 Eligible for Rehire: 001 Yes 002 Yes, but against company policy
(Circle only one) 003 No 004 Review required before rehiring

8 REASON FOR LEAVING
(Circle Only One)

101 Discharged (or Company Terminated Lease)
106 Laid Off (or Lease Suspended)
112 Leave of Absence
127 Retired
133 Resigned/Quit (or Driver Terminated Lease)
189 Other _____

9 STATUS
(Circle All That Apply)

202 Company Driver
207 Lease Driver (Employee of Independent Contractor)
213 Owner/Operator
228 Trip Leaser
230 Student/Trainee
234 Casual Driver
239 Other _____

10 DRIVER'S EXPERIENCE
(Circle All That Apply)

303 Local
311 Mountain Driving
327 Over the Road
332 Single Driver
333 Driver Trainer
351 1st Driver of a Team
352 2nd Driver of a Team
355 Freight Handling
399 Other _____

11 EQUIPMENT OPERATED
(Circle All That Apply)

504 Automobile
505 Auto Transporter
511 Bus
518 Double Trailer
523 Driveway/Towaway
527 Dry Box
529 Dump Truck
532 Flat Bed
534 Pick Up or Hot Shot
540 Refrigerated
542 Specialized Trailer
544 Specialized Truck/Toter
547 Straight Truck
552 Tank Truck
557 Taxi
562 Triple Trailer
573 Van
581 Winch
599 Other _____

12 LOADS HAULED
(Circle All That Apply)

707 Bulk Commodity
712 Container
713 Empty Trailer
714 Gen. Commodity
716 Electronics
718 Hanging Meat
720 Hazardous Material
725 Household Goods
729 Livestock
730 Lumber
731 Machinery
733 Mobile Homes
735 Motor Vehicles
762 Oversized Loads
763 Parcels
750 Passengers
764 Pipe
769 Refrigerated
773 Steel
799 Other _____

13 WORK RECORD
(Circle All That Apply)

901 Satisfactory
902 Superior
903 Outstanding
911 Complaints
913 Cargo Loss
917 Equipment Loss
915 Falsified Employ. Application
924 Late Pick Up/Delivery
926 Log Violation
928 No Show
929 Failed To Report Accident
931 Quit Under Dispatch
933 Quit/Dismissed During Training and/or Orientation
957 Unauthorized Equip. Use
959 Unauthorized Passenger
961 Unauthorized Use of Company Funds
935 Company Policy Violation
938 Unsatisfactory Safety Record
944 Personal Contact Requested
999 Other _____

QUIT UNDER LOAD/ABANDONMENT
(Circle Only One)

960 Co. Terminal - With Notice
961 Auth. Location - With Notice
952 Co. Terminal - W/O Notice
953 Unauth. Location - W/O Notice
954 Left Vehicle With Team Driver
955 Unauth. Location - With Notice
956 Auth. Location - W/O Notice

14 Prepared By _____ Date _____

15 Entered By _____ Date _____

16 This space for changes only.
TO CHANGE THIS RECORD

Record Identification Number _____

Signature approving change _____

Title _____

Appendix B
Alternative Termination Record Form
Work Record Section

[redacted] Form
Alternative Work Record Section

	Except- ional	Meets Expect- ation	Does not meet expect- ation	Not Observed
Job Knowledge & Skills				
Performs pre/post trip truck inspections				
Demonstrates good technical driving skill				
Is courteous & attentive to other traffic				
Demonstrates skill in adverse weather				
Maintains appropriate schedules				
Maintains logs				
Maintains equipment in good working order				
Maintains cargo responsibly				
Complies with state and federal hauling regulations				
Customer Relations				
Courteous and attentive to customer concerns				
Cooperative in scheduling adjustments				
Deals with emergencies effectively				
Communicates effectively with customers				
Personal Qualities				
Recognizes individual differences among customers				
Courteous with supervisors				
Courteous with co-workers				
Responsive to directions				
Communicates effectively with supervisors				
Attendance				
Personal appearance				